GenesLove.Me: A Model-based Web-application for Direct-to-consumer Genetic Tests

José Fabián Reyes Román[1,2], Carlos Iñiguez-Jarrín[1,3] and Óscar Pastor López[1]
[1]*Research Center on Software Production Methods (PROS), DSIC, Universitat Politècnica de València, Camino Vera s/n. 46022, Valencia, Spain*
[2]*Department of Engineering Sciences, Universidad Central del Este (UCE), Ave. Francisco Alberto Caamaño Deñó, 21000, San Pedro de Macorís, Dominican Republic*
[3]*Departamento de Informática y Ciencias de la Computación, Escuela Politécnica Nacional, Ladrón de Guevara E11-253, Quito, Ecuador*

Keywords: GLM, Direct-to-consumer Genetic Tests, Genetic Test, BPMN, Conceptual Models.

Abstract: The objective of this work was to enhance personalized medicine through the development and implementation of *Genomic Information Systems* (GeIS). For this, a web application called "*GenesLove.Me*" (GLM) was developed to provide direct-to-consumer genetic tests (DCGT). This paper focuses on the development of the business processes (BPMN) and a conceptual model (CM) for GLM, designed to analyze and improve the processes involved in this type of service and provide a model-based platform to manage genetic diagnoses in a *scalable*, *secure* and *reliable* way. Software Engineering (SE) approaches applied to the genomic context play a key role in the advancement of *personalized* and *precision medicine*.

1 INTRODUCTION

The current availability of direct-to-consumer genetic tests (DCGT) has a great number of advantages for the genomic domain, making it easier for end-users to access early genetic-origin diseases diagnosis services.

Romeo-Malanda (Romeo-Malanda, 2009) defines *"direct-to-consumer genetic analysis"* as a term which is used to describe analytic services offered to detect polymorphism and health-related genetic variations. Although this type of analysis is available through direct sales systems in pharmacies or other health care bodies, the Internet has become the main selling channel for direct-to-consumer genetic analyses. The usual procedure is to take a biological sample at home and send it to an analysis laboratory. The findings (results) of the analysis are communicated to the client by telephone, mail or electronic mail, or through secure access to an Internet portal (UNESCO, 2004).

The heterogeneity and dispersion of the data sources represent a great challenge, known today as *"genomic chaos"* (León et al., 2016). As the management of genomic repositories offers many benefits to the biomedical community, it is necessary to carry out studies and analyses to support the implementation of mechanisms to improve aspects related to data *integrity*, *consistency* and *homogeneity* (Reyes Román, 2014). In this way it will be possible to generate more reliable results for end-users by guaranteeing true *Precision Medicine* (PM).

The genomic data repositories (e.g., NCBI (NCBI Resource Coordinators, 2013), OMIM (Hamosh et al., 2005) and Ensembl (Cunningham et al., 2015) contain a very extensive set of information capable of being analyzed to extract concise data and generate results with greater accuracy. The efficient use of advances in genomic research allows the patient to be treated in a more direct way, which is reflected in results such as:"*better health*" and "*quality of life*".

In order to treat the data that will be used in the proposed application, we implemented the SILE (*Search-Identification-Load-Exploitation*) methodology (Reyes Román and Pastor López, 2016) which was conceived in the *PROS's Research Center* in order to improve selective loading processes of our *Human Genome Database* (HGDB).

Our aim in this work was twofold:

- To provide "*GenesLove.Me*" (GLM) as a web application designed to generate direct-to-consumer genetic tests (DCGT) supported by Business

Román, J., Iñiguez-Jarrín, C. and López, Ó.
GenesLove.Me: A Model-based Web-application for Direct-to-consumer Genetic Tests.
DOI: 10.5220/0006340201330143
In *Proceedings of the 12th International Conference on Evaluation of Novel Approaches to Software Engineering (ENASE 2017)*, pages 133-143
ISBN: 978-989-758-250-9

Process modeling (BPMN) (Object Management Group, 1997) and conceptual modeling (MC) (Olivé, 2007) techniques to study and analyze the essential elements of the processes involved in the genomic domain and improve the development of *Genomic Information Systems* (GeIS).

- To apply *Software Engineering* (SE) knowledge to improve *Precision Medicine* (PM). Ensuring PM for end-users involves implementing tools for studying, treating, exploring and exploding the genomic data generated and stored for several years to generate genetic diagnosis. The current DCGT demand was the reason for studying and analyzing the processes and stakeholders involved in this domain. The genetic tests are the result of the progress in the genomic environment with the aim of exploiting PM.

The genomic domain requires methodologies and modeling techniques capable of integrating innovative ideas into data management, process improvement and the inclusion of quality standards.

The paper is divided into the following sections: Section 2 reviews the present state of the art. Section 3 describes BPMN applied to genetic testing. Section 4 contains the representation of the domain by a conceptual model. Section 5 describes a case study with the GLM application, and Section 6 summarizes the lessons learned and outlines future work.

2 RELATED WORK

Bioinformatics now play an important role in contributing advances to the *medical* and *technological* sector. Genetic testing reveals existing knowledge about *"genes"* and *"variations"* in the genomic domain, which is used to diagnose diseases of genetic origin in order to prevent or treat them. This brings precision medicine closer to end-users (i.e., clients / patients).

The study of genomics (i.e., *data repositories*, *genetic variations*, *diseases*, *treatments*, etc.) is constantly growing and is increasingly seeking to ensure the application of precision medicine. DNA sequencing began in 1977 and since then software tools have been developed for its analysis. Thanks to NGS Technologies (Mardis, 2010), it is now possible to manipulate files (e.g., VCF) in order to generate genetic diagnoses in a more agile and efficient way (Reyes Román, 2014).

PM is a way of treating patients that allows doctors to identify an illness and select the treatment most likely to help the patient according to a genetic concept of the disease in question (this is why it has also been called *Personalized Medicine*) (Aguilar Cartagena, 2015).

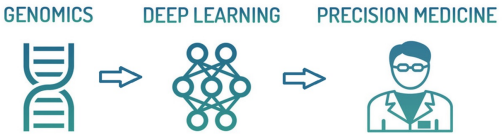

Figure 1: From Genomics to Precision Medicine.

As shown in Figure 1, this approach is based on a detailed knowledge of the genomic domain and on the information derived from the large amount of data generated in recent years. This information isconstantly growing as research provides more and more findings every day. Fowler et. al. (Jiménez, 2014) describe the advances in genomics that will provide practical information on diseases, those who are most likely to suffer them, and how most successful treatments can be applied, thus reducing the uncertainty and stress of patients and their families (Jiménez, 2014), (Instituto Nacional del Cáncer, 2015).

The advantages of genetic tests are innumerable and allow us to identify mutations or alterations in the genes and are of great use and interest in clinical / personalized medicine and the early diagnosis of diseases (Fisiotraining Córdoba, 2016), (Grupo RETO Hermosillo, 2016). Bioinformatics supports the tasks of: *management of biological databases*; *metabolic processes* and *population genetics*; *artificial intelligence* and others. The exploitation of genomics is of interest to various branches of research, such as: a) sequence analysis, b) genome annotation, c) evolutionary computational biology, d) gene expression analysis, e) protein expression analysis, f) cancer mutation analysis, g) comparative genomics, h) modeling of biological systems, i) protein-protein coupling, etc. (Franco et al., 2008)

By 2008 there were around 1,200 genetic tests available around the world (Fisiotraining Córdoba, 2016), (Grupo RETO Hermosillo, 2016), but they had some limitations (e.g., *data management*, *genome sequencing*, etc.) and their cost was quite high. For this reason, companies were interested in reducing costs and providing services to end-users in the comfort of their own homes. Technological advances played a fundamental role in the genomic environment, since the introduction of the NGS for sequencing samples made it possible to obtain sequences more quickly and cheaply (Metzker, 2010), (Voelkerding et al., 2009).

23andMe (23andMe, 2016a), in this sector. This American company offers a wide range of services. The type of information obtained from genetic samples is oriented to: *genetic history* (ancestors) and

personal health (risk of diseases), and is presented mostly in probabilistic terms (23andMe, 2016b). In the same way, in Spain companies of this type have emerged (e.g., *Genotest*[1] or *IMEGEN*[2]), all with the aim of providing genetic tests to end-users, simply and in the form of providing a diagnosis that allows end-users to take reactive or corrective actions (e.g., *prevention* and *treatment*) to improve their quality of life.

Genetic tests contain a large amount of sensitive information, so that before offering these services it is necessary to evaluate all the elements involved, such as: *ethical*, *moral*, *legal*, etc. It is important to be aware that the studies carried out must always become prevention instruments oriented to helping end-users. Companies within this business context treat sensitive information, so it is necessary to implement security (*high quality*) and data protection mechanisms, in addition to including all relevant legal measures. Spanish legislation requires these services to be provided under the following laws:

1. *Organic Law 15/1999, of December 13, on the Protection of Personal Data (Consolidated text dated March 5, 2011)* (BOE, 1999).
2. *Periods of Conservation of Personal Data in Biomedical Research* (BOE, 2011).
3. *General Health Law 86* (BOE, 2015).

For the implementation of DCGTs it is necessary to consider the aforementioned elements. In future years we will continue to improve our understanding of the human genome.

3 BPMN: DIRECT-TO-CONSUMER (DTC) GENETIC TEST

The GemBiosoft company is a spin-off of the Universitat Politècnica de València (UPV), founded in 2010.

The main objective of this company is to define the Conceptual Model of the Human Genome (CMHG) (Reyes Román et al., 2016), (Reyes Román et al., 2016) to obtain a precise schema to *manage*, *integrate* and *consolidate* the large amount of genomic data in continuous growth within the genomic domain. To achieve this objective, Gembiosoft has extensive experience in Model Driven-development (MDD) (Pastor et al., 2008) and an interdisciplinary team of people *-engineers and Ph.D.s-* trained to implement solutions aimed at i+D companies applying information technologies in the bioinformatics and health fields. GemBiosoft's collaborators include: PROS Research Center (UPV), PrincipiaTech and IMEGEN (*Instituo de Medicina Genómica*).

3.1 Case Description

GemBiosoft has a web application called *"GenesLove.Me"* which offers direct genetic testing to the consumer. The information provided by the genetic tests is accessible online to all users without prior registration (anonymous users). Figure 2 depicts the general use case of the interaction between actors and the web application, i.e. the application's functionality. For example, non-registered users of the web application are able to consult all information related to the diagnosis of rare diseases of genetic origin, their characteristics, treatment, tutorials and videos of the way in which the process is performed.

Access security in GenesLove.Me is controlled by profiles. Users can access GLM under 3 profiles: (1) *clients* (patients), (2) *provider* and (3) *administrator*. An authenticated user with a certain access profile is authorized to carry out the operations corresponding to the access profile.

(1) *Clients (patients)*: Users with this profile are able to contract the services offered by selecting the services (direct-to-customer genetic test) they are interested in and then paying the fee. The user is able to monitor the notifications and messages related to the diagnoses, besides consulting the histories of all the studies and treatments carried out and updating the information associated with his profile.

(2) *Supplier*: Users with this profile are able to generate notifications about the change of status in the treatment of samples. After receiving the genetic sample, the user activates the sample by entering its code number. They can then track the sample until the sequence file is generated. They can also update their profiles and consult all the activated samples (in progress and finalized).

(3) *Administrator*: A user with administration privileges performs administration and maintenance tasks of the web application, such as: a) publishing online results (the administrator uploads the resulting diagnoses from the analysis performed on samples sequenced by the *VarSearch*[3] tool. The application automatically notifies the user when his/her results have been published); b) publishing advertisements; c) publishing new diagnostic services to diag-

[1] http://www.trkgenetics.com/genotest

[2] https://www.imegen.es/

[3] VarSearch is a web application that seeks variation (or variant) files in a genomic database based on a Conceptual Model of the Human Genome (CMHG).

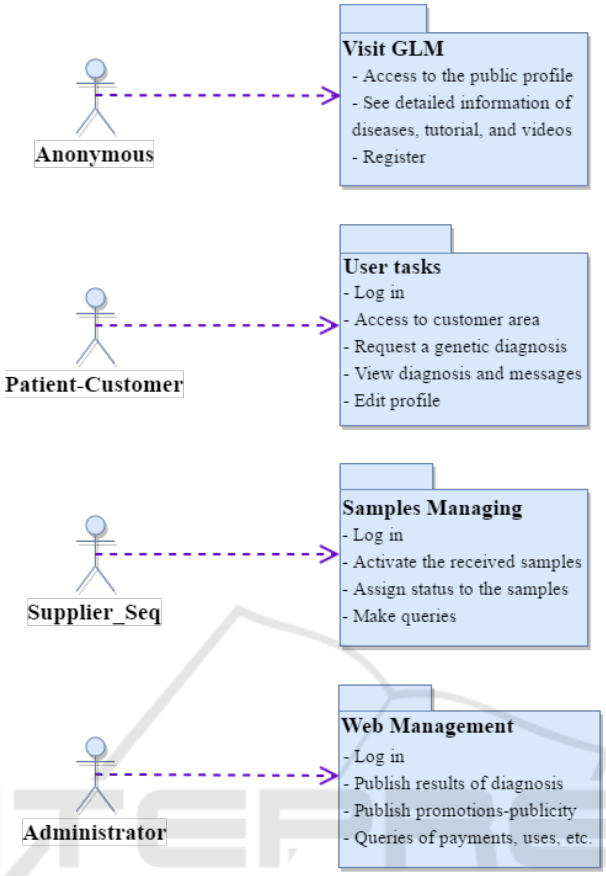

Figure 2: General view of the system.

nose new diseases; d) consulting payment reports and the application usage report (custom time period).

3.2 Genomic Diagnosis Process

Genetic tests are currently offered with the aim of detecting a person's predisposition to contracting a disease of hereditary origin (U. S. National Library of Medicine, 2017). The bioinformatics domain seeks to provide the necessary mechanisms and means to generate genetic diagnoses that allow the end-users (*patients*) to obtain these results to facilitate a personalized prevention treatment.

Figure 3 shows a BPMN diagram describing the genetic diagnosis process (from the end-user's service request until he/she receives his/her genetic test report). In this process, the three actors/users specified in Section 3.1 are involved: 1) The client (patient) who requests the service to determine whether or not he / she has a disease of genetic origin; 2) The company, in this case GemBiosoft, which is in charge of managing and performing the Genomic Diagnosis; and 3) the Suppliers, who in this case prepare the file containing the reference of the patient involved in the genetic test.

The general process begins when the end-user (patient) enters the web application and requests the genetic analysis (t1: task 1). The company (GemBiosoft) processes this request and proceeds to send the sample container to the client (t2). When the client receives the container, he must activate it by registering its identifier in the web application (t3), then place the sample in the container and send it back to the company (t4). Upon receipt of the sample, the company confirms that it meets the necessary requirements for the study and notifies the customer of its receipt (t5). The next step is to determine the supplier who will be responsible for sequencing the samples and send him the sample (t6). The selected supplier receives the sample and notifies its reception to the company (t7). Sequence preparation is initiated through the sequencing technology used by the supplier (t8). The supplier sends the resulting sequence of the sample (file) to the company (t9). The company confirms its reception to the supplier and proceeds to analyze the sequenced sample as part of the genetic diagnosis (t10). The definitive diagnosis report (t11) is then generated. The company (in this case the ad-

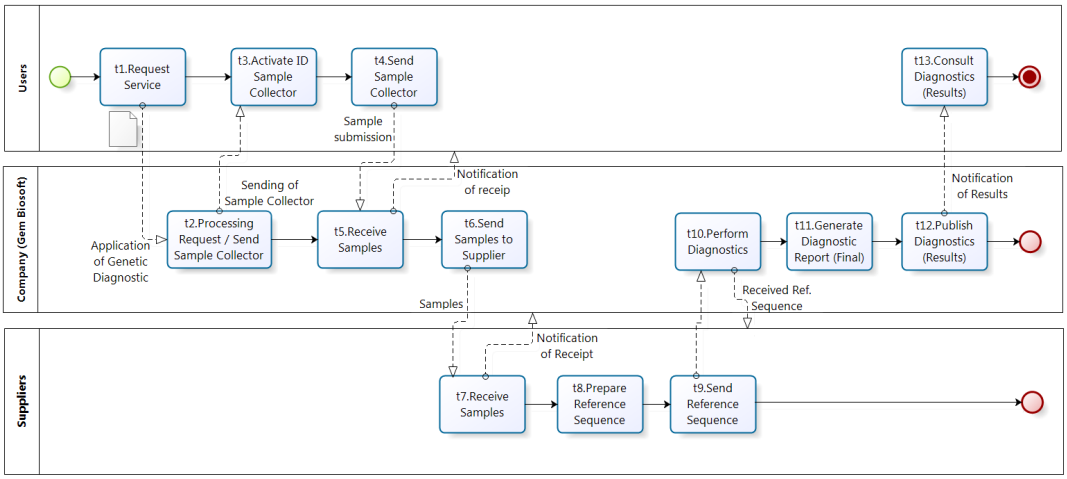

Figure 3: Genetic diagnosis process.

ministrator/user) proceeds to publish the genetic diagnosis (result) in the web application and the end-users are automatically notified of the results (t12). To end the process, the end-user accesses the web application to obtain the diagnosis and make any queries (results) (t13).

The BPMN (*Business Process Model and Notation*) gives companies the ability to understand their internal business procedures in graphical notation and the ability to communicate these procedures in a standard way (Chinosi and Trombetta, 2012). Through the model shown in Figure 3, it facilitates the understanding of commercial collaboration and transactions between organizations. In this figure we can see the interactions between end-user, company and suppliers (Reyes Román, 2014). The company is interested in providing a web application that allows end-users to obtain a quality genetic test in a simple way that aids the treatment and prevention of diseases of genetic origin.

4 REPRESENTATION OF THE DOMAIN: CONCEPTUAL MODEL

It is widely accepted that applying conceptual models facilitates the understanding of complex domains (like genetics) (Reyes R et al., 2016). In our case we used this approach to define a model representing the characteristics and the processes of direct-to-consumer genetic tests (DCGT).

One of the leading benefits of CM is that it accurately represents the relevant concepts of the analyzed domain (Reyes Román et al., 2016). After performing an initial analysis of the problem domain, the next step is to design a domain representation in the form of a CM. Our CM (Figure 4) evolved with the new discoveries made in the field of genomics (Reyes Román et al., 2016) in order to improve data processing to ensure effective precision medicine (PM). We can thus see how CM gives positive support to the knowledge in which precision medicine plays a key role (Reyes Román, 2014). It is important to highlight that the advantage of CM for representing this domain is that it eases the integration of new knowledge into the model (Reyes Román et al., 2016).

After an analysis of the requirements required for this work, important decisions were taken to arrive at an adequate representation of the basic and essential concepts in the understanding of the domain under study. Figure 4 presents the conceptual model proposed, which can be classified into three main parts: a) *Stakeholders*, b) *Genetic diagnostics* and c) *Sales management*.

The first part of our conceptual model represents all the participants involved in the web application. To represent the *"Users"* class of the inherit type we have the *"Administrator"*, *"Supplier_Seq"*, *"Patient Client"* and *"Anonymous"* classes respectively

The DCGTs are initialized when the end-user (*"patient customer"*) accesses the application. If the interaction does not have user credentials, it remains an *"anonymous"* user, otherwise, it becomes a client. The *"client customer"* contracts one or more services, which cover one or more diseases, represented in the model by the *"Services"* and *"Diseases"* classes, respectively.

Diseases are studied using the SILE methodology (Reyes Román and Pastor López, 2016), and when complete, the study is offered as a service. The end-user adds the services of interest to his shopping cart (represented by the *"Shopping Cart"* class) and then

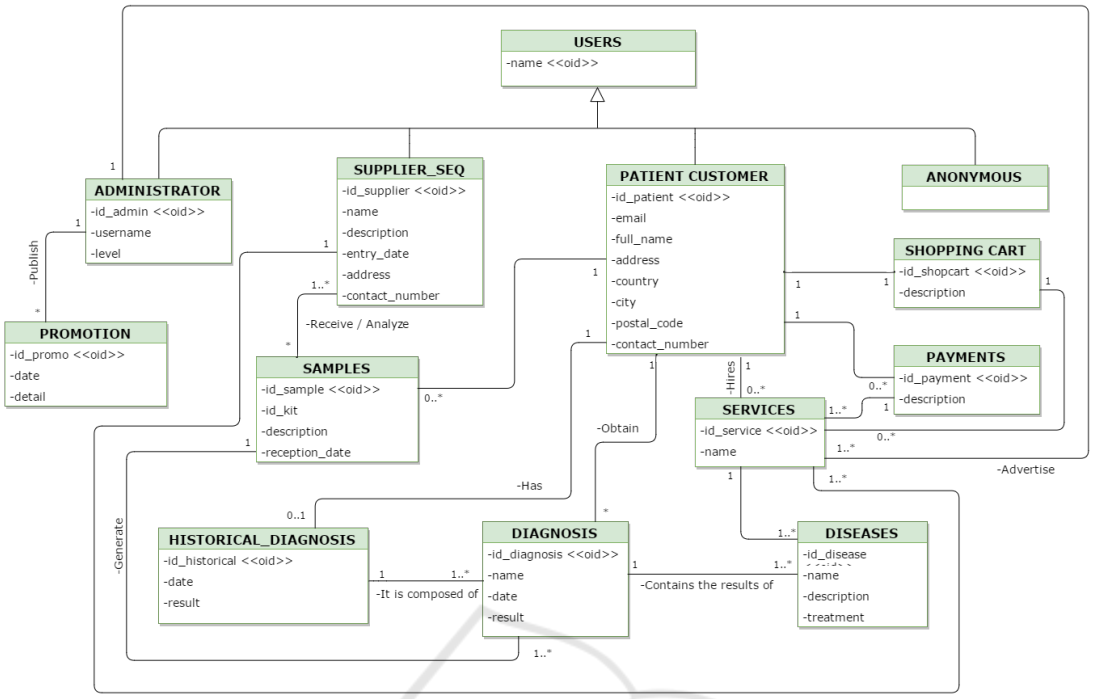

Figure 4: GLM conceptual model.

performs the payment process. For this, all the billing information, shipping address and acceptance of the purchase contract (which explains the rights and duties of the end-user and the company) is entered in the CM through the *"Payments"* class.

The management of genetic diagnoses is represented in the conceptual model through the *"Diagnostic"* and *"Historical diagnosis"* classes. The *"Diagnosis"* class shows the results obtained after the analysis of the sequence and a list of the diseases contracted. The user has the ability to consult his history associated with all the diagnoses requested by the application.

The diagnosis of genetic data begins when the company receives the file containing the sample sequenced by the supplier (represented by *"Supplier_Seq"* class). The application administrator is responsible for publishing the diagnosis results on the web and automatically notifies the end-user of the availability of his report. Other activities of the administration consist of the publication of advertising related to the implementation of new services and control of sales of the application.

Through our CM we incorporate genetic data currently used in the PM, achieving a conceptual representation that meets the needs of the bioinformatics domain. As we mentioned above, this model aims to improve the conceptual definition of the treatment related to genomic diagnosis, and thus leave a conceptual framework for further improvements.

5 CASE STUDY: GenesLove.Me SERVICE FOR END-USERS

In this section, we describe the design and implementation of *GenesLove.Me*[4] (GLM), an online web application based on the business process described in Section 3 of this work. The application becomes the point of interaction between three actors: the *clients* of genetic tests, the *company* that performs the genetic diagnosis and the *suppliers* who sequence the genetic samples. The main objective is to operate the DCGT service through an accessible online and easy-to-use platform for the actors involved in the process.

The platform offers a wide range of genetic tests on its main page (Figure 5), from which users are able to contract one or several genetic testing services (e.g., *lactose intolerance*, *androgenetic alopecia*, *alcohol sensitivity*, etc.) according to their needs. The company in charge of genetic analysis and diagnosis manages all the service orders received and monitors the change of order status throughout the service delivery process. For instance, the status of the order changes to *"KIT SENT TO SUPPLIER"* when the company (in charge of the analysis and diagno-

[4]http://www.geneslove.me

Figure 5: GenesLove.Me's home web page showing the available direct-to-customer genetic tests.

sis) sends the genetic sample to the suppliers responsible for sequencing the genetic samples and once the genetic samples have been sequenced and received from the supplier, the order is assigned *"ANALYSING SAMPLE"* status. Each status change is notified to the end-user via email and can be further visualized on the web user interface.

GLM is a web application implemented under a client-server architecture as is shown in Figure 6. The client side, the user's browser (e.g., *Firefox* (Mozilla, 2017), *Chrome* (Google, 2017), I*nternet Explorer* (Microsoft, 2017), etc.) serves as the interaction point between the user and application. The end-user interacts with an easy-to-use graphical web interface to request genetic testing services available in the application. The server side is hosted on the Internet and contains: a) the Apache 2.2 web server (The Apache Software Foundation, 2017) with the application logic implemented with PHP (W3schools, 2017) programming language and b) the data stored in the MySQL5.5 (Oracle Corporation, 2017) database engine.

The GLM design allows customers to access the online product range (i.e., *clinical tests*) from anywhere and at any time. In the same way, the administrator users in charge of site management and internal business tasks can access the GLM through an Internet connection.

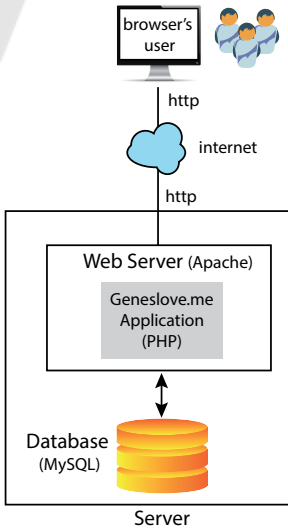

Figure 6: GenesLove.Me architecture.

GLM is implemented under Prestashop[5], an open source eCommerce CMS platform, which facilitates the implementation of customized solutions oriented to the marketing of products framed in a simple purchase-sale process. The platform incorporates both modules and website templates in order to provide, respectively, specific functionality and graphic style customized according to the business's needs. There is a great variety of free and commercial modules[6] and website templates created by web developers available on the Internet to be down- loaded and used.

The default Prestashop download package available on the official site includes modules of basic functionality (e.g., *customers*, *products*, *orders*, etc.) which are sufficient to create and manage a basic e-commerce platform. However, Prestashop allows complex functionality modules to be incorporated to tailor sites to particular needs. For instance, it is possible to install a module that deals with the complexity associated with the credit card payment process.

GLM takes full advantage of the range of functionalities offered by a CMS e-commerce platform such as Prestashop. Indeed, GLM offers a variety of on- line genetic tests, keeps the customer database and orders the genetic test sales process thanks to the available payment methods, such as *bank transfer*, *credit card* payments through *PayPal* (Jackson, 2012) or payments by *electronic check*.

The GLM client module represents the database of registered clients interested in the genetic tests. The ordering module lists the requests made by the customers, which consists of a reference code useful for order tracking, the name of the customer, the total value, the payment method and the current status of the order within the business process. The application's administrator manages the orders and updates the state of the orders. Updating the order state means a status change that automatically notifies the end user via e-mail. GLM facilitates the management of orders through the sequence of 13 states shown in Figure 7. Every state order change generates an email notification to keep customers informed of the processing of their orders. It is important to note that GLM includes interaction with *VarSearch* (see task 10 and 11 of Figure 3), an application developed by *PROS Research Center* to automatically identify the relevant information contained in the genomic databases and directly related to the genetic variations of the sequenced sample. *VarSearch* relies heavily on a *Conceptual Model of the Human Genome* (CMHG), which makes integration of external genomic databases feasible. However, due to the large amount of information available, the data loaded in *VarSearch* are the result of a selective loading process (Reyes Román and Pastor López, 2016) where the selected data correspond to the relevant information on the disease to be analyzed.

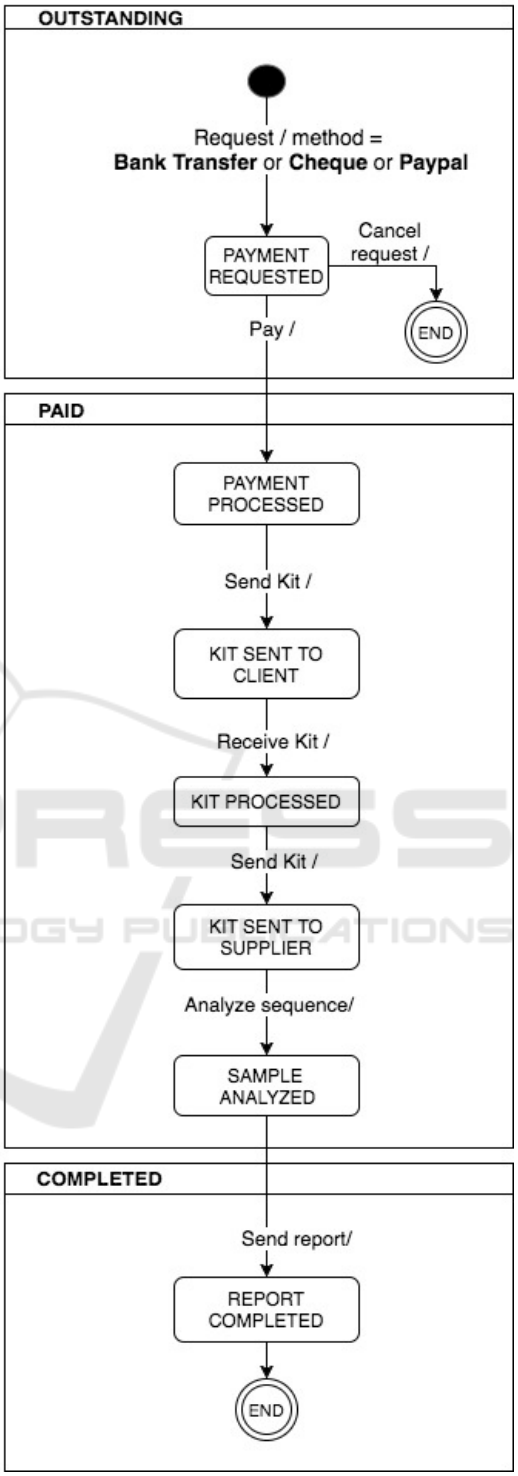

Figure 7: GLM's states sequence to manage orders.

[5]https://www.prestashop.com

[6](http://addons.prestashop.com/en/2modules)

The GLM user administrator has access to the administration panel (*back-office*) to manage the entire business process as well as the security and functionality of the site. The management screen has relevant information and direct access to the configuration and maintenance zones. For example, a strategically visible part of the main management screen displays information on the sales indicators and useful information for the management of the business accounting. By accessing the access security module, it is possible to configure profiles and access permissions for all users. It is also possible to manage the modules of *users*, *products*, *orders*, as well as the rules for discounts on purchases and customer service (e.g., *forums*, *e-mail notifications*). Thanks to the web platform on which the solution is implemented, site management can be carried out from anywhere and at any time.

Each module installed adds a functionality that supports to a greater or lesser degree the main reason of the platform: buying and selling products. The buying-and-selling process establishes the buyer-seller relationship through a simple process consisting of three states: *Pending*, *Paid* and *Sent*. "Pending" indicates that the user has requested a product, stored it in the shopping basket but still has to generate the payment. The *"Paid"* status indicates that the user has registered the payment for any of the options configured and available (e.g., *bank transfer*) and *"Sent"* indicates that the seller has shipped the product and is in the process of delivery to the customer.

5.1 Process Validation

In order to validate the process proposed in Section 3, test cases were performed with the implemented solution. The validation scenario consisted of a group of five (5) users, who made requests for genetic testing for *"lactose intolerance"*. To begin the process, each user involved in the case study authorized the procedure through an *"informed consent"* (Reyes Román, 2014), (de Galicia, 2001) which becomes a legal support that establishes the rights and obligations of the service offered and its expected scope.

6 LESSONS LEARNED AND FUTURE WORK

This paper describes a study and analysis of the implementation of a web application to facilitate DCGT, which informs end-users on their predisposition to suffer certain genetically based illnesses. Through the development of our web application *"GenesLove.Me"* we seek to provide end-users with a genomic diagnosis in a secure and reliable way.

The use of BPMN and Conceptual Modeling-based approaches for this type of service aids the understanding of the participants in the processes in the genomic domain and improves the processes involved. Through the proposed models we were able to evaluate different points:

1. *Process evaluation*: the processes involved in the direct-to-consumer genetic tests (DCGT) were studied, and a business model was developed to define the structures and relationships with the stakeholders.
2. *Improvement in data management*: a conceptual model (CM) was proposed for the definition of the relevant concepts in the domain with the objective of guaranteeing reliable and personalized medicine.

The validation of -*GenesLove.Me*- included the study of four disorders of genetic origin: *Alcohol Sensitivity*, *Androgenic Alopecia*, *Lactose Intolerance*, and *Dupuytren*. The research process was carried out using the SILE (*Search-Identification-Load-Exploitation*) Methodology, and after obtaining the variations/genes associated with the diagnoses, we made the selective loading of the data to our HGDB (*Human Genome Database*).

This web application was developed to provide end-users with an early genetic diagnosis through an attractive and easy-to-use distribution channel. Bioinformatics is a domain that is constantly evolving, and with the application of conceptual models we can extend our genomic knowledge and conceptual representation accurately and simply.

Future research work will focus on three main goals:

- The study and treatment of new diseases of genetic origin and continue expanding the list of illnesses available in the web application.
- To extend our Conceptual Model of the Human Genome (CMHG) by integrating new genomic information , to improve the generation of genetic diagnoses.
- Implementation of data management mechanisms to enhance the quality of personalized medicine.

ACKNOWLEDGEMENTS

The author thanks to the members of the PROS Center's Genome group for fruitful discussions. This work was supported by the Ministry of Higher

Education, Science and Technology (MESCyT) of Santo Domingo, Dominican Republic and the Secretaría Nacional de Educación, Ciencia y Tecnología (SENESCYT) and the Escuela Politécnica Nacional de Ecuador. The project also had the support of the Generalitat Valenciana through Project IDEO (PROM- ETEOII/2014/039) and the Spanish Ministry of Science and Innovation through Project DataME (ref: TIN2016-80811-P).

The author thanks Alberto García S., Mercedes R. Fernández Alcalá, Vicente Martínez Perelló and Verónica Burriel Coll for their collaboration with this this project.

REFERENCES

23andMe (2016a). 23andMe. https://www.23andme.com/. [Online; accessed 19-December-2016].

23andMe (2016b). How it works. https://www.23andme.com/howitworks/. [Online; accessed 19-December-2016].

Aguilar Cartagena, A. (2015). MEDICINA PERSONALIZADA, MEDICINA DE PRECISIÓN, ¿CUÁN LEJOS ESTAMOS DE LA PERFECCIÓN? *Carcinos*, 5:1–2.

BOE, A. e. B. O. d. E. (1999). Disposiciones Generales.

BOE, A. e. B. O. d. E. (2011). Ley 14/2007 de Investigación Biomédica.

BOE, A. e. B. O. d. E. (2015). Ley 14/1986 General de Sanidad.

Chinosi, M. and Trombetta, A. (2012). BPMN: An introduction to the standard. *Computer Standards and Interfaces*, 34(1):124–134.

Cunningham, F., Amode, M. R., Barrell, D., Beal, K., Billis, K., Brent, S., Carvalho-Silva, D., Clapham, P., Coates, G., Fitzgerald, S., Gil, L., Girón, C. G., Gordon, L., Hourlier, T., Hunt, S. E., Janacek, S. H., Johnson, N., Juettemann, T., Kähäri, A. K., Keenan, S., Martin, F. J., Maurel, T., McLaren, W., Murphy, D. N., Nag, R., Overduin, B., Parker, A., Patricio, M., Perry, E., Pignatelli, M., Riat, H. S., Sheppard, D., Taylor, K., Thormann, A., Vullo, A., Wilder, S. P., Zadissa, A., Aken, B. L., Birney, E., Harrow, J., Kinsella, R., Muffato, M., Ruffier, M., Searle, S. M. J., Spudich, G., Trevanion, S. J., Yates, A., Zerbino, D. R., and Flicek, P. (2015). Ensembl 2015. *Nucleic Acids Research*, 43(D1):D662–D669.

de Galicia, C. A. (2001). Ley 3/2001, reguladora del consentimiento informado y de la historia clínica de los pacientes.

Fisiotraining Córdoba (2016). Test Genéticos. http://www.fisiotraining.com/cordoba/analisisgeneticos_quees.htm. [Online; accessed 19-December-2016].

Franco, M. L., Cediel, J. F., and Payán, C. (2008). Breve historia de la bioinformática. *Colombia Médica*, 39:117–120.

Google (2017). Personalizar Chrome.

Grupo RETO Hermosillo, A. (2016). El cáncer de mama. http://gruporetohermosilloac.com/index.php. [Online; accessed 19-December-2016].

Hamosh, A., Scott, A. F., Amberger, J. S., Bocchini, C. A., and McKusick, V. A. (2005). Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic Acids Research*, 33(DATABASE ISS.):D514–D517.

Instituto Nacional del Cáncer (2015). Qué es la medicina de precisión en el tratamiento del cáncer. https://www.cancer.gov/espanol/cancer/tratamiento/tipos/medicina-de-precision. [Online; accessed 19-December-2016].

Jackson, E. M. (2012). *The PayPal Wars: Battles with eBay, the Media, the Mafia, and the Rest of Planet Earth*. WND Books, 2nd edition.

Jiménez, N. (2014). Una medicina nueva , más inteligente y menos invasiva. http://www.lifescienceslab.com.

León, A., Reyes, J., Burriel, V., and Valverde, F. (2016). Data quality problems when integrating genomic information. In *International Conference on Conceptual Modeling*, pages 173–182. Springer.

Mardis, E. R. (2010). The $1,000 genome, the $100,000 analysis? *Genome Medicine*, 2(11):84.

Metzker, M. L. (2010). Sequencing technologies - the next generation. *Nature reviews. Genetics*, 11(1):31–46.

Microsoft (2017). Internet Explorer - Microsoft Download Center.

Mozilla (2017). Firefox.

NCBI Resource Coordinators (2013). Database resources of the National Center for Biotechnology Information. *Nucleic Acids Research*, 41(D1):D8–D20.

Object Management Group (1997). Business Process Model and Notation.

Olivé, A. (2007). *Conceptual Modeling of Information Systems*. Springer Berlin Heidelberg, Berlin, Heidelberg, 1 edition.

Oracle Corporation (2017). MySQL :: MySQL 5.7 Reference Manual.

Pastor, O., España, S., Panach, J. I., and Aquino, N. (2008). Model-driven development. *Informatik-Spektrum*, 31(5):394–407.

Reyes R, J. F., Pastor, O., Valverde, F., and Roldán, D. (2016). Including haplotypes treatment in a Genomic Information Systems Management. In *Proceedings of the XIX Ibero-American Conference on Software Engineering (CIbSE 2016)*, pages 27–29, Quito, Ecuador.

Reyes Román, J. F. (2014). Integración de haplotipos al modelo conceptual del genoma humano utilizando la metodología sile. http://hdl.handle.net/10251/43776.

Reyes Román, J. F., Pastor, Ó., Casamayor, J. C., and Valverde, F. (2016). Applying conceptual modeling to better understand the human genome. In Comyn-Wattiau, I., Tanaka, K., Song, I.-Y., Yamamoto, S., and Saeki, M., editors, *Conceptual Modeling: 35th International Conference, ER 2016, Gifu, Japan, November 14-17, 2016, Proceedings*, pages 404–412. Springer International Publishing, Cham.

Reyes Román, J. F., Pastor Lopez, O., Valverde, F., and Roldán, D. (2016). How to deal with Haplotype data : An Extension to the Conceptual Schema of the Human Genome. *CLEI ELECTRONIC JOURNAL*, 19(2).

Reyes Román, J. F. and Pastor López, Ó. P. (2016). Use of GeIS for Early Diagnosis of Alcohol Sensitivity. In *Proceedings of the 9th International Joint Conference on Biomedical Engineering Systems and Technologies*, pages 284–289.

Romeo-Malanda, S. (2009). Análisis genéticos directos al consumidor: cuestiones éticas y jurídicas. http://www.institutoroche.es/legalactualidad/85/analisis. [Online; accessed 19-December-2016].

The Apache Software Foundation (2017). Welcome! - The Apache HTTP Server Project.

U. S. National Library of Medicine (2017). What is genetic testing? - Genetics Home Reference.

UNESCO (2004). Declaración Internacional sobre los Datos Genéticos Humanos. http://unesdoc.unesco.org/images/0013/001361/136112so.pdf. [Online; accessed 19-December-2016].

Voelkerding, K. V., Dames, S. A., and Durtschi, J. D. (2009). Next-generation sequencing:from basic research to diagnostics. *Clinical Chemistry*, 55(4):641–658.

W3schools (2017). PHP 5 Tutorial.